



Probabilistic Analysis in Linear Programming

Ron Shamir

Statistical Science, Vol. 8, No. 1, Report from the Committee on Applied and Theoretical Statistics of the National Research Council on Probability and Algorithms. (Feb., 1993), pp. 57-64.

Stable URL:

<http://links.jstor.org/sici?sici=0883-4237%28199302%298%3A1%3C57%3APAILP%3E2.0.CO%3B2-G>

Statistical Science is currently published by Institute of Mathematical Statistics.

Your use of the JSTOR archive indicates your acceptance of JSTOR's Terms and Conditions of Use, available at <http://www.jstor.org/about/terms.html>. JSTOR's Terms and Conditions of Use provides, in part, that unless you have obtained prior permission, you may not download an entire issue of a journal or multiple copies of articles, and you may use content in the JSTOR archive only for your personal, non-commercial use.

Please contact the publisher regarding any further use of this work. Publisher contact information may be obtained at <http://www.jstor.org/journals/ims.html>.

Each copy of any part of a JSTOR transmission must contain the same copyright notice that appears on the screen or printed page of such transmission.

JSTOR is an independent not-for-profit organization dedicated to and preserving a digital archive of scholarly journals. For more information regarding JSTOR, please contact support@jstor.org.

Probabilistic Analysis in Linear Programming

Ron Shamir

Abstract. The main results on probabilistic analysis of the simplex method and on randomized algorithms for linear programming are reviewed briefly.

Key words and phrases: Average-case analysis, linear programming, probabilistic analysis, randomized algorithms, the simplex method.

1. THE LINEAR PROGRAMMING PROBLEM

The *linear programming* (LP) problem calls for minimizing a linear objective function subject to linear constraints. Every such problem can be presented in the following form:

$$(P) \quad \min \{c^T x \mid Ax \geq b, x \geq 0\}.$$

Here the vectors $c^T = (c_1, \dots, c_d)$, $b^T = (b_1, \dots, b_m)$ and the $m \times d$ matrix A constitute the input, and $x = (x_1, \dots, x_d)$ is the vector of variables to be optimized. Usually, one assumes that all input numbers are integers or rationals. The goal is to find an optimal x if such exists, and if not, to determine that the problem is unbounded (i.e., the minimum is $-\infty$) or infeasible (i.e., there is no x satisfying all the constraints).

LP is one of the fundamental problems in operations research and computer science. It was formulated in the 1940s by George Dantzig, who also proposed a method for solving it. Since then, the problem has been studied in thousands of scientific papers and hundreds of books, and interest in the problem has stayed on a very high level. One reason for this is the wide applicability of LP: it is used to model practical, real-life problems in economics, industry, communications, military and numerous other areas. It is perhaps the most widely used optimization model in the world, and except data structures, perhaps the largest single use of computer resources (see Lovász, 1980.) Another reason is that LP problems with special structure arise in various theoretical and practical areas. Yet another reason is that some fundamental properties of LP and its algorithms are still not fully understood. One of these is the efficiency of the simplex method, which will be discussed here.

The method that was proposed by Dantzig for solving LP is the simplex method (e.g., Dantzig, 1963). To describe it geometrically, we need some terminology:

A point in \mathbb{R}^d is *feasible* if it satisfies all the constraints. The set of feasible points,

$$X = \{x \in \mathbb{R}^d \mid Ax \geq b, x \geq 0\},$$

is a polyhedron, that is, the intersection of a finite set of half-spaces. Each half-space is of the form $\{x \mid A_i \cdot x \geq b_i\}$, and its corresponding hyperplane is $\{x \mid A_i \cdot x = b_i\}$. ($A_i \cdot$ denotes the i th row in A .) For simplicity, we assume that the problem is *nondegenerate*; that is, each set of d hyperplanes meets at a distinct single point. In that case, if $X \neq \emptyset$, the dimension of X is d , and each feasible point in which exactly d of the inequalities are satisfied as equalities is a *vertex* of X . Vertices are *adjacent* if they are connected by an edge of X , or, equivalently, if their defining sets of equalities have $d - 1$ identical elements. It is not hard to see that if there is an optimal solution, its value will be obtained at some vertex of X . Assuming nondegeneracy, the set of feasible points lying on one defining hyperplane is (if nonempty) of dimension $d - 1$ and is called a *facet* of the polyhedron.

The simplex method starts from a feasible vertex and moves in each iteration to an adjacent vertex until eventually either an optimal vertex is reached or the algorithm recognizes at some vertex that the problem is unbounded. The move from one vertex to an adjacent one is called a *pivot*, and it can be realized mathematically by a simple manipulation of the constraints matrix. Usually, the next vertex is chosen so that the objective function value will improve after the pivot step. Even so, there is still a lot of freedom in choosing the next vertex. The specific rule of this choice is called a *pivoting rule*. Numerous pivoting rules have been devised for the simplex method, each giving rise to a different variant of the method. To obtain the initial feasible vertex, one can set up another LP problem (called the *phase I* problem) to which there is always a known feasible vertex and apply the simplex method to that problem. This problem is set up in such a way that it is bounded, and when its optimal solution is reached, one either has a starting feasible vertex for the original problem or knows that that problem is infeasible. The scope of this article does not allow us

Ron Shamir is Associate Professor, School of Mathematics, Tel Aviv University, Tel Aviv, 69978 Israel.

to get into more detail. We refer the interested reader to the (very incomplete) reference list. Also, several parenthetical comments on duality that are made in the sequel can be ignored by the reader who is not familiar with it. They are needed for accuracy but not for understanding the essence of the results.

Within a few years of its introduction, LP had become a central – perhaps the central – paradigm of operations research. The simplex method had proved itself to be extremely useful and highly efficient in practice. However, by the early 1970s it became clear that the simplex is not – as far as we know – a theoretically efficient algorithm: the complexity of the simplex method was shown to be finite but not polynomial in the input length. In fact, on specially structured “bad” examples, many simplex variants were shown to require a number of pivots exponential in the problem dimensions. Hence, as far as we know, the simplex is *not* a good algorithm from the theoretical point of view. On the other hand, Monte Carlo studies and vast practical experience with real-life problems show that the typical number of pivots is very low: it appears to be a linear or slightly superlinear function of the dimensions. How can we explain the huge gap between the provably bad worst-case bounds and the efficiency of the method in practice? A natural approach is to apply probabilistic analysis: assume that the input data are distributed according to some prespecified distribution and show that *on the average* the number of pivots is small. Another approach is to equip the algorithm with a randomization mechanism, and show that for every input, averaged over the possible actions of the algorithm, the number of pivots is small.

This paper gives a brief survey of various results from probabilistic analysis of the simplex method. The scope of the paper forces us to omit many details, to say nothing of proofs. We shall be able to give only the flavor of the results and the different approaches, in the hope of kindling the reader’s curiosity. The interested reader may find more detailed surveys in Borgwardt (1987), Todd (1991), Megiddo (1987) and in Shamir (1987), from which an earlier version of this paper originated. Much more on the context of the problem can be found in the references thereof.

2. PROBABILISTIC ANALYSIS

In order to perform probabilistic analysis, one has to specify the simplex variant used, the initialization (phase I) procedure, the form of LP used and the probabilistic assumption on the distribution of problem inputs. Given those parameters, we denote by $\rho(n, d)$ [or $\rho(m, d)$] the average number of pivots for problems with n inequalities in d variables [or m inequalities in d nonnegative variables].

All the probabilistic models discussed here generate problems that are nondegenerate with probability one.

To avoid being too technical, we do not describe the conditions that guarantee nondegeneracy in each model but simply assume throughout that all the models deal only with (primal and dual) nondegenerate problems.

2.1 Parametric Simplex Variants

The variant of the simplex that has been proved most amenable to probabilistic analysis is the parametric objective simplex algorithm due to Gass and Saaty (1955). This variant – explicitly or in disguise – plays a central role in most of the studies discussed below. Let us sketch the parametric simplex method and introduce some terminology that is needed later. Suppose we have a LP with two objective functions, the original objective c and a *co-objective* c' . Given an optimal solution to problem (P) with objective c' , the goal is to solve problem (P) with the original objective c . The problem can be presented as one with a parametric objective:

$$(P(\lambda)) \quad \min\{\lambda c^T x + (1 - \lambda)c'^T x \mid Ax \geq b, x \geq 0\}.$$

Here λ is a scalar parameter, a solution for $P(0)$ is known, and we want to solve problem $P(1)$. In the parametric simplex algorithm, the value of λ is gradually increased. A new problem is obtained for each λ , but for a certain range of λ values the optimal vertex remains the same. When the current optimal vertex ceases to be optimal, a pivot step is performed, and the algorithm moves (along an edge) to a new vertex that is optimal for larger λ . The process terminates when the value $\lambda = 1$ is reached or when the solution to the problem becomes unbounded at some critical value $\lambda < 1$, in which case problem $P(1)$ is unbounded. Under our blanket assumption of nondegeneracy, the set of optimal points visited by the algorithm is a path of vertices and edges. This set,

$$\{x \in \mathbb{R}^d \mid x \text{ is optimal for some } P(\lambda), 0 \leq \lambda \leq 1\},$$

is called the *efficient path*. Hence the number of vertices along that path equals the number of pivot steps performed by the parametric simplex algorithm. Another set of points can be obtained by the same algorithm when the value of λ decreases from 0 to $-\infty$. The solution for $\lambda = -\infty$ corresponds to the solution for maximizing (rather than minimizing) $c^T x$. The set of solutions for $-\infty < \lambda \leq 1$ is called the *co-optimal path*.

A similar idea can be applied to the following problem, in which b and c are parametrized simultaneously:

$$(II(\lambda)) \quad \min\{\lambda c^T x + (1 - \lambda)c'^T x \mid Ax \geq \lambda b + (1 - \lambda)b', x \geq 0\}.$$

The *parametric self-dual (PSD)* simplex algorithm, which was introduced by Dantzig (1963), starts from an optimal solution to $II(0)$. It increases λ gradually

and performs a sequence of (primal or dual) pivots that yield solutions for $\Pi(\lambda)$ for increasing values of λ . Note that with an initial choice of $c^T = (1, 1, \dots, 1)$ and $-b^T = (1, 1, \dots, 1)$, $x = 0$ is an optimal solution for $\Pi(0)$, so this algorithm avoids the need for two simplex phases.

The convenience in analyzing the parametric simplex is due to its property that, given a vertex of the feasible set, one can determine directly if the simplex path passes through that vertex. "Directly" means by an algebraic closed-form condition depending on the vertex and the input data, so the pivots need not be performed.

2.2 Borgwardt's Results

The first major probabilistic analysis results were due to Karl Heinz Borgwardt (1977a, b; 1978). He analyzed problems of the form

$$\min\{c^T x \mid Ax \leq e\},$$

where A is of dimension $n \times d$ and $e^T = (1, 1, \dots, 1) \in \mathbb{R}^n$. Note that such problems are always feasible, since $x = 0$ is feasible. The basic probabilistic model he used requires that each of the n vectors constituting the rows of A and of the vector c assume independently the same spherically symmetric distribution on $\mathbb{R}^d - \{0\}$. In other words, the distribution of each such vector is invariant under any rotation around the origin. One can visualize such a distribution as a "cloud" around the origin, with the density of the cloud at each point proportional to the probability of choosing the vector from the origin to that point. In a spherically symmetric distribution, the density at each point on any sphere centered at the origin is the same, but spheres of different radii may have different densities. Hence the radial part of the distribution is the only part that may vary among such distributions.

Borgwardt used the parametric simplex. Borgwardt (1978) showed that when c' is also chosen randomly and independently according to the same distribution, the expected number of pivots along the efficient path for d fixed and $n \rightarrow \infty$ is asymptotically bounded by

$$\text{const} \cdot d^2 n^{1/(d-1)}.$$

A crucial step in Borgwardt's analysis was to restate the problem in terms of the polar set X^* of the feasible set X , which is also a polyhedron (Grünbaum, 1967). The number of vertices along the efficient path equals the number of faces of X^* intersected by a two-dimensional cone spanned by the two objectives c and c' . Borgwardt then estimates the expectation of that number by an integral expression, which eventually leads to the results.

Borgwardt (1977a, b) also obtained interesting asymptotic upper and lower bounds for several specific spherically symmetric distributions. For example,

when all the mass of the distribution is on the unit sphere, he obtained a lower bound of $d^{1.5} n^{1/(d-1)}$ on the expected number of pivots. He also showed that, whenever d is fixed and $n \rightarrow \infty$, $\rho(n, d) \rightarrow \infty$ for every distribution with bounded support, and that there exist distributions for which $\rho(n, d)$ grows slower than any polynomial in n .

Later, Borgwardt (1981, 1982) obtained a nonasymptotic bound on the expected length of such an efficient path, namely,

$$\text{const} \cdot d^3 n^{1/(d-1)}.$$

In order to obtain a result for a complete (two-phased) simplex method, Borgwardt defined a new algorithm. It generates a sequence of subproblems with increasing numbers of variables and uses the parametric algorithm to solve each subproblem. Borgwardt (1982) proves a polynomial expected bound for that algorithm. This was the first proof that the average number of pivots of any complete simplex variant is polynomial. The bound for Borgwardt's complete algorithm and model was subsequently improved (Borgwardt, 1987) and now stands at

$$(B) \quad \rho(n, d) \leq \text{const} \cdot (d + 1)^4 n^{1/(d-1)}.$$

Recently, Borgwardt (private communication, 1992) has improved the asymptotic analysis for the complete algorithm (i.e., phase I plus phase II) to $\text{const} \cdot d^{2.5} n^{1/(d-1)}$.

Because of the importance of Borgwardt's work, this section concludes by noting three problems in his model, which may be improved upon in the future: first, the model generates only feasible problems, for which a feasible point is known in advance. Second, the complete algorithm that was "tailored" to facilitate the analysis can be used only on problems whose form fits the model. [For some progress on these two problems, see Borgwardt (1990).] Third, its phase I solves a sequence of $n - 1$ problems to get a feasible point for phase II, although such a point is known a priori. For experimental studies on the model described above, see Borgwardt et al. (1990).

2.3 Asymptotic Results: Smale and Others

Smale (1983a, b) investigated linear programs of the form (P) . The probabilistic model in Smale (1983a) requires that the vector $(c^T, -b^T)$ assume a spherically symmetric distribution in $\mathbb{R}^{d+m} - \{0\}$ and A assumes a spherically symmetric distribution in $\mathbb{R}^{m \times d} - \{0\}$.

The variant analyzed was the PSD simplex. Smale used a theory developed by Eaves and Scarf (1976) that implies a presentation of the PSD algorithm for problem $\Pi(\lambda)$ as a set of piecewise-linear equations. Using this presentation, Smale (1983a) showed how to express the probability that a vertex is in the PSD-path in terms of the volume of a corresponding cone. A series of estimates for the volume of the special cone

structure generated for LP eventually yielded the result

$$(S) \quad \rho(m, d) \leq c(d) \cdot (1 + \log(m + 1))^{d(d+1)}.$$

In particular, this result implies that when d is fixed, $\rho(m, d)$ grows with m more slowly than any fixed positive power of m . The dependence on d is, however, exponential.

Smale's model seems to have a certain advantage over Borgwardt's, because it uses a more practical variant and there is no restriction on the form of the linear programs it can solve. Also, the model generates both feasible and infeasible problems. However, the asymptotic results assume that d is fixed and $m \rightarrow \infty$; in that situation most problems generated under this model become infeasible, so it is not clear how to interpret such results. We shall return to this point below.

Following Smale, Blair (1986) showed that essentially the same asymptotic result can be obtained by a simpler argument, under somewhat weaker probabilistic assumptions. Smale's estimate (S) is meaningful only when d is fixed and $m \rightarrow \infty$. Blair approached that case directly. One way to restate his observation is that when $m \gg d$, under Smale's probabilistic model, most of the constraints are "dominated" by other constraints and cannot participate in vertices of the PSD path. (Intuitively, a constraint is dominated if another constraint separates it from the origin in the positive orthant. We omit the formal definition of domination here.) By obtaining an upper bound of the expected number of undominated constraints, he proved that for fixed d and $m \rightarrow \infty$,

$$\rho(m, d) \leq c(d) \cdot (\log m)^{d(d+1) \log(d+1) + d}.$$

The exponent of $\log m$ is higher than in (S), but the result is essentially the same. Namely, for fixed d , $\rho(m, d)$ grows with m slower than any polynomial. Blair also showed that the same result holds for two other simplex variants.

Adler, Karp and Shamir (1986) investigated several probabilistic models that generalize Smale's. Their most general model requires for LP of the form (P), in addition to some nondegeneracy assumptions, that the data distribution be invariant with respect to reversing the signs of any subset of the rows of the matrix (A, b). They define a family of algorithms that proceed according to a "constraint-by-constraint" (CBC) principle. This principle requires that vertices that satisfy as equality the k th constraint may be considered only if the subproblem defined by the $k - 1$ preceding constraints has been shown to be feasible. [A dual form of the CBC method is closely related to the phase I given in Borgwardt (1982).] This facilitates the detection of infeasibility of a problem at a relatively early

stage in order to exploit the high probability that a problem is infeasible under this model when $m \gg d$.

Using the CBC principle, the authors show that even if *full enumeration* of both feasible and infeasible intersection points of d hyperplanes is done in every subproblem,

$$\rho(m, d) \leq \text{const} \cdot 2^{5d},$$

independent of m . In particular, when d is fixed and $m \rightarrow \infty$, the expected number of steps is bounded by a *constant*. By using more efficient CBC algorithms and stronger probabilistic assumptions, the upper bound is reduced to a smaller function of d , but it still remains exponential. The authors also show that these results hold for a broad family of simplex variants.

Although the bound obtained above is better than in Smale (1983a), it is not an improvement on Smale's, because the variant that he analyzed is not a CBC algorithm. Using the approach developed in Smale (1983a), Megiddo (1986a) improved the analysis for Smale's model to

$$\rho(m, d) \leq \bar{c}(d).$$

Furthermore, Megiddo showed that $\rho(m, d)$ decreases to the limit for any fixed d . However $\bar{c}(d)$ is superexponential in d .

The results of Blair (1986) and Adler, Karp and Shamir (1986) emphasized a considerable drawback of the asymptotic analyses of sign-invariant (and spherically symmetric) models: Blair showed that Smale's result can be explained in terms of the small expected number of vertices of the feasible set, without really using the properties of the specific parametric simplex variant. Adler, Karp and Shamir showed that upper bounds for $\rho(m, d)$ that depend on d only can be obtained by *enumeration* procedures, without using any property of the simplex method. Both of these results indicate that those asymptotic bounds reflect the properties of the *probabilistic model* rather than those of the simplex method. Later studies, as we shall soon see, overcame this drawback by obtaining results that are meaningful for all d and n .

Later, Smale (1983b) extended his results to a more general probabilistic model. The main addition in that model is the replacement of spherical symmetry with invariance of the probability distribution under coordinate permutations of \mathbf{R}^d . The results of Adler, Karp and Shamir (1986) and Megiddo (1986a) do not apply to the more general model. Blair's result does apply to Smale's second model, and it indicates a similar drawback in interpreting asymptotic results under it.

Recently, Portnoy (1991) used probabilistic techniques to analyze the number of pivots in the parametric simplex algorithm. His motivation was to study a statistical regression problem, which led to a somewhat

specific LP, but one in which random variation in the constraints arise naturally. For fixed dimension, when the number of constraints n tends to infinity, he showed that the number of pivots is $O(n \log n)$ in probability.

2.4 Adler, Haimovich: Sign-Invariant Model for Phase II

Adler (1983) and Haimovich (1983) studied independently the expected number of pivots along paths generated by some parametric simplex variants. We present their results with respect to the form $P(\lambda)$. The probabilistic model they used assumes *sign invariance*: the input distribution must be invariant under changing the signs of any subset rows and columns of

$$\begin{pmatrix} c^T & 0 \\ c'^T & 0 \\ A & b \end{pmatrix}.$$

(Changing the sign of row i simply replaces the inequality $A_{i \cdot} x \leq b_i$ by $A_{i \cdot} x \geq b_i$. Changing the sign of column i is equivalent to changing the constraint $x_i \geq 0$ to $x_i \leq 0$. In other words, the hyperplane does not change; rather, the other half-space that is supported by it is chosen.) In addition, a certain nondegeneracy condition is required. This model generalizes Smale's first model, since every spherically symmetric distribution satisfies the sign-invariance condition. Certain discrete distributions and certain problems with sparse matrices also fall under this model.

The main advantage of this model is that it facilitates analysis in terms of simple counting arguments. Under this model, Adler and Haimovich showed that the average number of pivots along a nonempty co-optimal path in a feasible LP is no more than

$$\min(m, d) + 1.$$

Similar results were obtained using other parameterizations. In particular, Adler showed the same bound for the PSD simplex with the sign-invariance assumptions extended to the additional right-hand side vector b' .

The Adler-Haimovich results were very significant but still less than a complete explanation of the good real-life behavior of the simplex method. These results show that the co-optimal path from a point that minimizes a random objective to a point maximizing that objective is on average short. One problem is that this analysis assumes that a phase II algorithm starts from a random point, and this assumption has not been proved so far for any known variant. Furthermore, the co-objective must be chosen independently of the particular problem. In known phase I algorithms the starting point is not random but is either fixed or dependent on the particular data. Such differences may change the results dramatically. There are striking examples from linear complementarity theory (Meg-

iddo, 1986b; Saigal, 1983) where a path starting from a random point has expected polynomial length and a path starting at a fixed point has expected exponential length.

2.5 The Quadratic Results

Toward the end of 1983, three independent investigations obtained an upper bound on $\rho(m, d)$ that is *quadratic* in $\min(m, d)$. Two of the studies, Todd (1986) and Adler and Megiddo (1985), analyzed the PSD simplex. The third, Adler, Karp and Shamir (1987), analyzed a parametric version of the CBC method. The probabilistic model was the same in all three investigations. It was essentially the sign-invariance model described in the previous section, with a stronger nondegeneracy assumption. It is still more general than the model in Smale (1983a).

The PSD simplex was analyzed under a specific choice of so-called lexicographic initialization vector. Todd analyzed that algorithm using tools from matroid theory, whereas Adler and Megiddo developed further Smale's approach. Adler, Karp and Shamir analyzed a parametric CBC algorithm with a lexicographic co-objective, building on the previous results and ideas of Adler and Haimovich.

After the completion of these three investigations, Megiddo (1985) observed that, although the parametric CBC algorithm and the PSD algorithm are in general quite different, their lexicographic versions execute exactly the same sequence of pivots. Thus, all three investigations were concerned with *the same simplex variant* and obtained the result

$$\rho(m, d) \leq \text{const} \cdot [\min(m, d)]^2.$$

Later, Adler and Megiddo (1985) obtained a quadratic lower bound on the average number of pivots. The probabilistic assumptions required for that result are stronger. They require that all coordinates of the data (A, b, c) are independent, identically distributed random variables with a common distribution symmetric about the origin. Their result was obtained by establishing bounds on the correlation between the probabilities for bases to be both primal and dual feasible simultaneously. That result, together with the upper bound above, establishes that, at least for the stronger model, $\rho(m, d)$ is in fact bounded between two quadratic functions of $\min(m, d)$. This also proves that there cannot exist a subquadratic upper bound for the weaker model.

Unlike the Adler-Haimovich results, the quadratic bound was not obtained under conditioning of feasibility. The sign-invariance model manifests the same behavior as Smale's spherically symmetric model; namely, as the ratio n/d grows, the proportion of feasible problems diminishes, and the meaning of averaging the num-

ber of pivots over all feasible and infeasible problems is put in question. However, using the fact that the quadratic result holds for all n and d , it was shown that for $n = 2d$ the expected number of pivots in a feasible problem under that model is at most $\text{const} \cdot d^{2.5}$.

3. RANDOMIZED ALGORITHMS

Suppose we allow a simplex algorithm to choose randomly among several possibilities at certain points along its computation. For example, the algorithm can randomly choose to which vertex to move next among the adjacent vertices that provide a better value to the objective. We can then compute the expected number of pivots required to solve a problem when the average is taken over the possible random choices. The expected number of pivots of such a randomized algorithm is the largest average number of pivots on any problem with the same size. (The maximum is over all problems of this size, and the average is over the internal randomizations performed by the algorithm.) Hence in this approach the algorithm is randomized, and the input data are not. In contrast, in the probabilistic analyses described in previous sections the algorithm is completely deterministic, and the randomization is over the problem set.

Randomized pivoting rules have been suggested in the past for LP; see Dantzig (1963). Recently, there were several important developments in the theory of randomized algorithms for LP. Dyer and Frieze (1989) and Clarkson (1991) described a randomized algorithm for LP in case the dimension d is fixed, improving the deterministic algorithm of Megiddo (1984) for this special case. Clarkson's algorithm has expected complexity $O(d^2n + d^{d/2+O(1)} \log n)$. Clarkson's method was also exploited to obtain a fast parallel algorithm for LP in fixed dimension (Alon and Megiddo, 1990) and generalized to convex programming (Adler and Shamir, 1989). Seidel (1990, 1991) described another very simple "dual simplex" randomized algorithm for LP in fixed dimension, of complexity $\Theta(d!n)$. This has been improved by Sharir and Welzl (1992) and extended to cover a fairly general class of convex programming problems, to a dual simplex algorithm with $O(n \cdot 2^d)$ expected number of pivots. A previous attempt at improving Seidel's algorithm is due to Welzl (1991). It seems to run very fast in practice but no analysis of its complexity is known yet.

In a remarkable recent development, Kalai (1992b) has given the first randomized simplex algorithm for general LP problems that is subexponential. His algorithm takes on the average

$$n^{C \cdot \sqrt{d \log d}}$$

pivots, where C is an absolute constant. Roughly speaking, his algorithm starts from a vertex v , then

reaches vertices on r different facets, randomly chooses one facet among them and works recursively on that facet to find the optimal vertex in it. A judicious choice of r (and a careful analysis) leads to the above bound.

Although this bound is not polynomial, this is a substantial improvement over the exponential upper bounds known previously. Kalai's work was preceded by important results on the diameter of polyhedra. [See Kalai (1991, 1992a) and Kalai and Kleitman (1992) and the excellent survey by Klee and Kleinschmidt (1987) for the context of this problem. In that context, Kalai has also shown that if one does not limit the computational effort per pivot step, $n^{\log d}$ pivot steps always suffice.] Shortly after Kalai's result, Matoušek, Sharir and Welzl (1992) showed that a careful analysis of the previous randomized simplex algorithm of Sharir and Welzl (1992) leads to the same expected complexity bound.

4. THE ROAD AHEAD

In retrospect, one can see two parallel lines of research on probabilistic simplex analysis. The first line of research, on the feasible rotation-symmetric model, was carried out single-handedly by Borgwardt. The second started with Smale's model and was later generalized to the sign-invariant model, leading finally to the quadratic results. In both models the analysis advanced in three stages:

1. asymptotic results for fixed dimension,
2. nonasymptotic results for phase II, starting from a vertex optimal with respect to a random objective, and
3. results for a complete simplex algorithm.

In both cases the average complexity and the "naturalness" of the variant had to be sacrificed to some extent between stages 2 and 3.

Although the gap between the good practical behavior of the simplex and its proven bad worst-case bounds has been explained in part by the probabilistic analyses, many related questions are still open. Both models have some undesirable properties that were mentioned above. Analyzing, or even merely defining, a widely acceptable model that generates "real-life" problems is a major challenge. Any analysis of the concentration of the number of pivots around the mean (e.g., the variance) will be new and interesting. More natural simplex variants, especially those used in practice, are also still awaiting analysis.

A short time after the quadratic results were obtained, Karmarkar (1984) announced a new algorithm for LP that is both theoretically and practically efficient. The interest of the research community shifted to this exciting development, and for a while some people believed that the simplex method had met its

match and was now obsolete. Now that most of the dust has settled, it looks as if both algorithms have their respective advantages, and each has problems that it solves faster. Moreover, usually it is not possible to tell a priori which algorithm will be faster on a given problem. (One of the wonderful side benefits of this development is that the existing codes for the simplex method have been improved by almost two orders of magnitude to make them competitive with the new challenger.) As is often the case, new solutions raise new problems, and one of them is related to our theme: the efficiency of Karmarkar's method in practice is much faster than what can be proved theoretically. Explaining this gap by probabilistic analysis is a new great challenge for the LP community.

In randomized algorithms for LP, a major challenge is now to find a randomized simplex algorithm that is polynomial or to prove that no such algorithm exists.

ACKNOWLEDGMENTS

This article was written while the author was a visitor at DIMACS and RUTCOR at Rutgers University, supported by AFOSR Grants 89-0512 and 90-0008 and by NSF Grant NSF-STC88-09648.

I thank Ilan Adler and Micha Sharir for their helpful comments on an early version of the manuscript. I also thank Dorit Hochbaum who suggested the idea of writing this paper.

REFERENCES

- ADLER, I. (1983). The expected number of pivots needed to solve parametric linear programs and the efficiency of the self-dual simplex method. Working Paper, Dept. Industrial Engineering and Operations Research, Univ. California, Berkeley.
- ADLER, I., KARP, R. M. and SHAMIR, R. (1986). A family of simplex variants solving an $m \times d$ linear program in expected number of pivots depending on d only. *Math. Oper. Res.* 11 570-590.
- ADLER, I., KARP, R. M. and SHAMIR, R. (1987). A simplex variant solving an $m \times d$ linear program in $O(\min(m^2, d^2))$ expected number of pivot steps. *J. Complexity* 3 372-387.
- ADLER, I. and MEGIDDO, N. (1985). A simplex algorithm whose average number of steps is bounded between two quadratic functions of the smaller dimension. *J. Assoc. Comput. Mach.* 32 871-895.
- ADLER, I. and SHAMIR, R. (1989). A randomized scheme for speeding up algorithms for linear and convex programming problems with high constraints-to-variables ratio. Technical Report 89-7, DIMACS Center, Rutgers Univ. *Math. Programming*. To appear.
- ALON, N. and MEGIDDO, N. (1990). Parallel linear programming in fixed dimension almost surely in constant time. In *Proceedings of the 31st IEEE Symposium on the Foundations of Computer Science* 574-582. IEEE Computer Society Press, Los Alamitos, CA.
- BLAIR, C. (1986). Random linear programs with many variables and few constraints. *Math. Programming* 34 62-71.
- BORGWARDT, K. H. (1977a). *Untersuchungen zur Asymptotik der mittleren Schrittzahl von Simplexverfahren in der Linearen Optimierung*. Ph.D. dissertation, Kaiserslautern Univ., Germany.
- BORGWARDT, K. H. (1977b). Untersuchungen zur Asymptotik der mittleren Schrittzahl von Simplexverfahren in der Linearen Optimierung. *Methods Oper. Res.* 28 332-345.
- BORGWARDT, K. H. (1978). Zum Rechenaufwand von Simplexverfahren. *Methods Oper. Res.* 31 83-97.
- BORGWARDT, K. H. (1981). The expected number of pivot steps required by a certain variant of the simplex method is polynomial. *Methods Oper. Res.* 43 35-41.
- BORGWARDT, K. H. (1982). The average number of pivot steps required by the simplex-method is polynomial. *Z. Oper. Res.* 26 157-177.
- BORGWARDT, K. H. (1987). *The Simplex Method: A Probabilistic Approach*. Springer, New York.
- BORGWARDT, K. H. (1990). Probabilistic analysis of the simplex method. In *Mathematical Developments Arising from Linear Programming* (J. C. Lagarias and M. J. Todd, eds.) 21-34. Amer. Math. Soc., Providence, RI.
- BORGWARDT, K. H., DAMM, R., DONIG, R. and JOAS, G. (1990). Empirical studies on the average efficiency of simplex variants under rotation symmetry. Report, Inst. Mathematik, Univ. Augsburg, Germany.
- CLARKSON, K. L. (1991). A Las Vegas algorithm for linear programming when the dimension is small. Draft manuscript, AT&T Bell Laboratories. [Earlier version with same title in *Proceedings of the 20th Annual Symposium on Theory of Computing* (1988) 452-456. ACM Press, New York.]
- DANTZIG, G. B. (1963). *Linear Programming and Extensions*. Princeton Univ. Press.
- DYER, M. E. and FRIEZE, A. M. (1989). A randomized algorithm for fixed-dimensional linear programming. *Math. Programming*. 44 202-212.
- EAVES, C. and SCARF, H. (1976). The solution of systems of piecewise linear equations. *Math. Oper. Res.* 1 1-27.
- GASS, S. I. and SAATY, E. L. (1955). The computational algorithm for the parametric objective function. *Naval Res. Logist.* 2 39-45.
- GRÜNBAUM, B. (1967). *Convex Polytopes*. Wiley, New York.
- HAIMOVICH, M. (1983). The simplex algorithm is very good! - On the expected number of pivot steps and related properties of random linear programs. Draft, Columbia Univ.
- KALAI, G. (1991). The diameter problem for convex polytopes and f -vector theory. In *The Victor Klee Festschrift* (P. Gritzman and B. Sturmfels, eds.) 387-411. Amer. Math. Soc., Providence, RI.
- KALAI, G. (1992a). Upper bounds for the diameter of graphs of convex polytopes. *Discrete Comput. Geom.* 8 363-372.
- KALAI, G. (1992b). A subexponential randomized simplex algorithm. In *Proceedings of the 24th Annual ACM Symposium on Theory of Computing* 479-487. ACM Press, New York.
- KALAI, G. and KLEITMAN, D. J. (1992). A quasi-polynomial bound for diameter of graphs of polyhedra. *Bull. Amer. Math. Soc.* 26 315-316.
- KARMARKAR, N. (1984). A new polynomial-time algorithm for linear programming. *Combinatorica* 4 373-395.
- KLEE, V. and KLEINSCHMIDT, P. (1987). The d -steps conjecture and its relatives. *Math. Oper. Res.* 12 718-755.
- LOVÁSZ, L. (1980). A new linear programming algorithm - better or worse than the simplex method? *Math. Intelligencer* 2 141-146.
- MATOUŠEK, J., SHARIR, M. and WELZL, E. (1992). A subexponential bound for linear programming. In *Proceedings of the 8th Annual Symposium on Computational Geometry* 1-8. ACM Press, New York.
- MEGIDDO, N. (1984). Linear programming in linear time when the dimension is fixed. *J. Assoc. Comput. Mach.* 31 114-127.

- MEGIDDO, N. (1985). A note on the generality of the self-dual algorithm with various starting points. *Methods Oper. Res.* **49** 271–275.
- MEGIDDO, N. (1986a). Improved asymptotic analysis of the average number of steps performed by the self dual simplex algorithm. *Math. Programming* **35** 140–172.
- MEGIDDO, N. (1986b). On the expected number of linear complementarity cones intersected by random and semi-random rays. *Math. Programming* **35** 225–235.
- MEGIDDO, N. (1987). On the complexity of linear programming. In *Advances in Economic Theory* (T. F. Bewley, ed.) 225–268. Cambridge Univ. Press.
- PORTNOY, S. (1991). Asymptotic behavior of the number of regression quantile breakpoints. *SIAM J. Sci. Statist. Comput.* **12** 867–883.
- SAIGAL, R. (1983). On some average results for linear complementarity problems. Manuscript, Dept. Industrial Engineering and Management Sciences, Northwestern Univ.
- SEIDEL, R. (1990). Linear programming and convex hulls made easy. In *Proceedings of the 6th Annual ACM Symposium on Computational Geometry* 211–215. ACM Press, New York.
- SEIDEL, R. (1991). Low dimensional linear programming and convex hulls made easy. *Discrete Comput. Geom.* **6** 423–434.
- SHAMIR, R. (1987). The efficiency of the simplex method: A survey. *Management Sci.* **33** 301–334.
- SHARIR, M. and WELZL, E. (1992). A combinatorial bound for linear programming and related problems. In *Proceedings of the 9th Symposium on the Theoretical Aspects of Computer Science. Lecture Notes in Comput. Sci.* **577** 569–579. Springer, New York.
- SMALE, S. (1983a). On the average speed of the simplex method of linear programming. *Math. Programming.* **27** 241–262.
- SMALE, S. (1983b). The problem of the average speed of the simplex method. In *Mathematical Programming: The State of the Art* (A. Bachem, M. Grötschel and B. Korte, eds.) 530–539. Springer, New York.
- TODD, M. J. (1986). Polynomial expected behavior of a pivoting algorithm for linear complementary and linear programming problems. *Math. Programming* **35** 173–192.
- TODD, M. J. (1991). Probabilistic models for linear programming. *Math. Oper. Res.* **16** 671–693.
- WELZL, E. (1991). Smallest enclosing disks (balls and ellipsoids). In *New Results and New Trends in Computer Science* (H. Maurer, ed.). *Lecture Notes in Comput. Sci.* **555** 359–370. Springer, New York.